Brownian Dynamics Simulation of Protein Solutions: Structural and Dynamical Properties

Paolo Mereghetti,[†‡*] Razif R. Gabdoulline,[§] and Rebecca C. Wade[†*]
[†]Heidelberg Institute for Theoretical Studies, Heidelberg, Germany; [‡]The Interdisciplinary Center for Scientific Computing, University of Heidelberg, Heidelburg, Germany; and [§]BIOBASE, Wolfenbuttel, Germany

ABSTRACT The study of solutions of biomacromolecules provides an important basis for understanding the behavior of many fundamental cellular processes, such as protein folding, self-assembly, biochemical reactions, and signal transduction. Here, we describe a Brownian dynamics simulation procedure and its validation for the study of the dynamic and structural properties of protein solutions. In the model used, the proteins are treated as atomically detailed rigid bodies moving in a continuum solvent. The protein-protein interaction forces are described by the sum of electrostatic interaction, electrostatic desolvation, nonpolar desolvation, and soft-core repulsion terms. The linearized Poisson-Boltzmann equation is solved to compute electrostatic terms. Simulations of homogeneous solutions of three different proteins with varying concentrations, pH, and ionic strength were performed. The results were compared to experimental data and theoretical values in terms of long-time self-diffusion coefficients, second virial coefficients, and structure factors. The results agree with the experimental trends and, in many cases, experimental values are reproduced quantitatively. There are no parameters specific to certain protein types in the interaction model, and hence the model should be applicable to the simulation of the behavior of mixtures of macromolecules in cell-like crowded environments.

INTRODUCTION

An important feature of living systems is the presence of highly concentrated media consisting of many macromolecules immersed in solvent. Biological molecules, which have evolved to work in crowded media, may exhibit different kinetic and thermodynamic properties under high concentration conditions than those observed at the low concentration conditions typically used in in vitro (1–3) experiments. Indeed, it has been shown that crowding leads to nonnegligible effects on several biologically important processes such as protein folding, macromolecular association, and biochemical reactions, as well as on the stability of protein structures (3–8). The study of the effects of crowders on biological processes and the understanding of the underlying molecular details is challenging due to the wide range of effects produced by crowding molecules (9). However, due to improved computing capacity, atomically detailed models can now be used to describe simple cell-like environments from a molecular and structural viewpoint (10–13).

We have developed a Brownian dynamics (BD) method for simulating many macromolecules (10^2–10^3) described as atomically detailed rigid bodies in a continuum solvent in a periodic box. The model is based on that implemented in the Simulation of Diffusional Association (SDA) software (14). The physical model for the intermolecular forces has been designed for general applicability (see Methods for details). In line with recent comments by Elcock (15), we have paid particular attention to the quantitative agreement of our results with experimental data.

BD simulations of three different, experimentally well-characterized, monodisperse protein solutions were carried out to validate the model and the simulation procedure. The proteins studied were hen egg-white lysozyme (HEWL), bovine pancreatic trypsin inhibitor (BPTI), and T4 lysozyme. The structural, dynamic, and thermodynamic aspects of these systems were studied. The self-diffusion coefficients, osmotic second virial coefficients, structure factors, and aggregation properties were analyzed. For all the systems studied, the experimental trends were reproduced and, in many cases, a quantitative agreement between experimental and computed values was achieved.

In Methods, we first describe the protein-protein interaction model for computing forces. Then, we describe how the analysis for comparison to experimental observables was performed. Finally, the structure preparation, system setup, and BD simulation details are given. Results and Discussion consists of three main subsections for the computation of B_{22} values, oligomeric states, and self-diffusion coefficients, respectively. Within each subsection, the three proteins are discussed separately. In the Conclusions, we provide an outlook on the future perspectives arising from this study.

METHODS

Interaction energies and forces

Forces are computed as finite-difference derivatives of the pairwise free energies of interaction between proteins. For each pair of proteins, the interaction free energy, ΔG, is defined as

Submitted May 19, 2010, and accepted for publication October 20, 2010.
*Correspondence: rebecca.wade@h-its.org or paolo.mereghetti@h-its.org
Editor: Ruth Nussinov.

0006-3495/10/12/3782/10 $2.00
doi: 10.1016/j.bpj.2010.10.035

$$\Delta G = \frac{1}{2}\sum_{i_2}\Phi_{el_1}(\mathbf{r}_{i_2})\cdot q_{i_2} + \frac{1}{2}\sum_{j_1}\Phi_{el_2}(\mathbf{r}_{j_1})\cdot q_{j_1} + \sum_{i_2}\Phi_{edesolv_1}(\mathbf{r}_{i_2})\cdot q_{i_2}^2 + \sum_{j_1}\Phi_{edesolv_2}(\mathbf{r}_{j_1})\cdot q_{j_1}^2 + \sum_{m_2}\Phi_{npdesolv_1}(\mathbf{r}_{m_2})\cdot SASA_{m_2} + \sum_{n_1}\Phi_{npdesolv_2}(\mathbf{r}_{n_1})\cdot SASA_{n_1} + \sum_{m_2}E_{softcore_1}(\mathbf{r}_{m_2}) + \sum_{n_1}E_{softcore_2}(\mathbf{r}_{n_1}). \quad (1)$$

For computational efficiency, all interaction potentials, Φ, are mapped onto cubic grids. A detailed description and parameterization of the first six terms of Eq. 1 can be found in the literature (16,17). Briefly, the first two terms in Eq. 1 are the interaction energies of the charges of one protein (q_{i_2}or q_{j_1}) with the electrostatic potential of the other protein (Φ_{el_1} or Φ_{el_2}). Charges were assigned using the effective charge approximation (17).

The third and fourth terms of Eq. 1 represent the electrostatic desolvation energy arising from the introduction of the low dielectric cavity of one protein in the presence of the charges of the other (17,18). It is computed as the interaction of the charges of one protein (q_{i_2}or q_{j_1}) with the electrostatic desolvation potential of the other protein ($\Phi_{edesolv_1}$ or $\Phi_{edesolv_2}$) (18), with parameterization as in Gabdoulline and Wade (16).

The fifth and sixth terms in Eq. 1 correspond to the nonpolar interactions due to the burial of the solvent-accessible surface areas of the surface atoms of one protein by the other protein. This includes a factor β for converting the calculated buried area to hydrophobic desolvation energy (16), which was set to $\beta = -0.018$ kcal mol^{-1} Å^{-2} in this work. This value gave the best correspondence with experimental results in all systems tested and is the range from -0.013 to -0.019 kcal mol^{-1} Å^{-2} used in Gabdoulline and Wade (16) for studies of the diffusional association of plastocyanin and cytochrome *f*.

The last two terms of Eq. 1 describe the soft-core repulsive potential introduced to avoid overlaps. The soft-core potential was modeled using the following inverse power function

$$E(\mathbf{r})_{softcore} = \gamma\sum_i \frac{1}{\left(\frac{a_i}{\sigma}\right)^{nexp} + \left|\mathbf{r}-\mathbf{r}_i\right|^{nexp}}, \quad (2)$$

where r_i is the center of atom i of radius a_i. The term

$$\left(\frac{a_i}{\sigma}\right)^{nexp}$$

removes the singularity at $|\mathbf{r}–\mathbf{r}_i| = 0$ and gives a smooth function. The value of σ can be tuned to vary the smoothness of the function, keeping its asymptotic behavior. The value γ is a parameter to rescale the magnitude of the function and *nexp* is the exponent to set the decay. These parameters were chosen to reproduce the pairwise radial distribution function in lysozyme simulations, which do not include the soft-core potential. In the latter, overlaps are avoided by checking every step for clashes and if a move would result in a clash, trying new moves until the proteins do not overlap (14,16,19). The parameters were set to $nexp = 6$, $\sigma = 3$ Å, and $\gamma = 64.0$ kcal/mol.

Because the electrostatic desolvation, nonpolar desolvation and soft-core potential terms are short range, the size of the corresponding grids can be kept small. In particular, for the smaller proteins studied here (BPTI and HEWL), the grid size was set to 60^3 Å^3, whereas for T4 lysozyme, the dimension of the desolvation grid was 80^3 Å^3 and that of the soft-core grid was set to 60^3 Å^3.

For the electrostatic potential, the grid dimensions were set to 80^3 Å^3 for all systems at ionic strengths (IS) ranging from 100 to 800 mM. At 5 mM IS (only used for HEWL), the size of the grid was increased to obtain an electrostatic potential at the border of ≤ 0.05 kcal/mol/*e*. Specifically, at pH 3 (net charge $+13e$), the grid size was set to 200^3 Å^3 whereas at pH values 6 and 9 (with corresponding net charges of $+9e$ and $+7e$, respectively), the grid was set to 150^3 Å^3.

The electrostatic potential grid was computed by solving the linearized Poisson-Boltzmann equation using UHBD (20). The protein and solvent relative dielectric constants were set to 2.0 and 78.4, respectively. The protein interior was defined by the van der Waals surface. All grids were computed using a 1.0 Å grid spacing.

ANALYSIS

Osmotic second virial coefficient, B_{22}

The B_{22} value was computed from the radial distribution function, $g(r)$, as (21,22)

$$B_{22} = -2\pi\int_0^\infty (g(r)-1)r^2dr, \quad (3)$$

where r is the distance between the centers of geometry of the proteins. We compared computed B_{22} values to experimental ones obtained using static light scattering, where the Rayleigh scattering equation (23) was used to derive second virial coefficients. Because the electrostatic interactions are computed with a grid-charge formalism in the simulations, interactions beyond the extent of the grid are neglected. We therefore computed an analytical long-range correction to the B_{22} value following Asthagiri et al. (24) and Hill (25) and considering each protein as a single Debye-Hückel sphere beyond the electrostatic potential grid. Further details regarding the computation of B_{22} values from simulations and the long-range correction can be found in the Supporting Material.

Oligomer analysis and clustering

An average fraction of oligomeric species was computed by recording the occurrence of the oligomeric states at each step of the simulation and then averaging over the total number of steps. The criteria used to define an oligomer and the details of the computation can be found in the Supporting Material.

Cluster analysis was carried out to find the most favorable orientations in each oligomeric species. Computational details are described in the Supporting Material.

Structure factors

Structure factors, $S(q)$, were computed by Fourier transformation of $g(r)$ (26).

Self-diffusion coefficient

The self-diffusion coefficient was obtained from the time evolution of the time-ensemble averaged $\langle\cdot\rangle_{te}$ mean-squared displacement computed as

$$\left\langle r^2(\tau)\right\rangle_{te} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{t^{end}-\tau+1}\sum_{t=0}^{tend-\tau}\left[\mathbf{r}_i(t+\tau)-\mathbf{r}_i(t)\right]^2, \quad (4)$$

where N is the total number of proteins, t^{end} is the total simulation time, and $\mathbf{r}_i(t)$ is the center of geometry of protein i at time t.

A linear fit was performed on a closed interval $[\tau_1, \tau_2]$, and D_s was computed from the slope using the Einstein relation (see the Supporting Material).

The observed reduction of the self-diffusion coefficient with increasing concentration was compared to two different theoretical models. The first model, derived by Han and Herzfeld (27), is based on scaled particle theory (28), whereas the second model was developed by Tokuyama and Oppenheim (29) in deriving the short- and long-time self-diffusion coefficients for concentrated monodisperse suspensions of hard spheres with hydrodynamic interactions. The description of these models can be found in the Supporting Material.

Protein structure preparation

The crystal structures of the proteins used in the simulations were downloaded from the Protein DataBank (30): 1HEL (31), 2HEX (32), and 1L87 (33) for HEWL, BPTI, and T4 lysozyme, respectively.

Polar hydrogens were added to the structures using H++ (34). Protonation states were set according to the specific pH based on experimental pK_a values. For HEWL, pK_a values at low IS were obtained from Abe et al. (35), whereas for IS values >100 mM, pK_a values from Demchuk and Wade (36) were used. Experimental pK_a values for T4 lysozyme were obtained from Kuhn et al. (37), and those for BPTI were taken from Demchuk and Wade (36). For HEWL, the net charge of differently protonated structures at different pH values was compared to experimental titration curves (38), obtaining a good agreement (see Fig. S1 in the Supporting Material).

Partial charges and radii were assigned to all the atoms from the OPLS force field (39).

The structures of the S44K and S44F mutants of T4 lysozyme were built using MODELER (40).

The physical properties of the proteins are given in Table S1 in the Supporting Material.

System setup and simulations

BD simulations were carried out using 512 proteins. Each system was subjected to 10 μs of simulation at 300 K, using periodic boundary conditions. Further details about the setup and simulations are described in the Supporting Material.

BD simulations were performed with SDAMM, a parallelized program based on the SDA software (14) capable of handling many proteins (10^3–10^4) treated as rigid bodies in atomic detail.

Parallelization was implemented using openMP (www.openmp.org), mainly parallelizing the loops for computation of the forces. A 1-μs simulation of 512 lysozyme molecules in a $\approx$110 nm^3 box required approximately one day on eight shared memory 2.4-MHz Opteron processors.

For comparison with experimental data, each of the analyzed properties (i.e., B_{22} and D_s values) compared was obtained from a separate simulation under conditions that reproduced the corresponding experimental conditions.

RESULTS AND DISCUSSION

Osmotic second virial coefficients (B_{22} values)

Osmotic second virial coefficients (B_{22}) appear as coefficients in the virial expansion of the osmotic pressure (21). A positive value of the B_{22} indicates an osmotic pressure larger than that for an ideal solution, which, at the molecular level, can be interpreted to reflect an average net repulsive interaction among the solute molecules. The opposite is true for a negative value of B_{22} (11,21,41–46). Although B_{22} values can be computed simply from consideration of pairwise interactions, our purpose here was to compare B_{22} values derived from experiments with values computed from the $g(r)$ at low concentrations ($\approx$10–20 g/L) as a means to assess our many-protein simulations, as has been done previously (11).

HEWL

Nine simulations were carried out for combinations of three different IS and three different pH values at a protein concentration of 10 g/L to reproduce the experimental conditions described in Velev et al. (22). At 100 mM IS, the computed B_{22} values are in good agreement with the experimental data (see Fig. 1). At 5 and 500 mM IS, the sign and trend of the B_{22} values are correct.

Fig. 1 shows the B_{22} values obtained with and without adding the analytical long-range correction (see Methods). This correction is based on the assumption that the proteins behave as Debye-Hückel spheres beyond the boundaries of

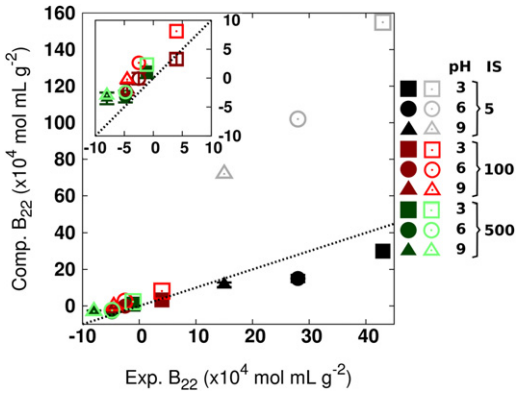

FIGURE 1 Plot of computed against experimental (22) B_{22} values for HEWL. (IS: 5 mM, *black*; 100 mM, *red*; 500 mM, *green*; pH 3, *squares*; pH 6, *circles*; and pH 9, *triangles*.) A reference line with gradient 1 is dashed. The computed values are given with and without the analytical long-range correction as open and solid symbols, respectively.

the electrostatic grid. The assumption is reasonable because, for all the systems we simulated, the standard deviation of the electrostatic potential at the grid boundaries is $<10\%$ of the mean value. This analytical correction strongly affects the computed B_{22} values at 5 mM IS, which become much higher than the experimental values from Velev et al. (22). However, more recent data from Asthagiri et al. (24) show that at very low ionic strength (7 mM), the B_{22} values should be much larger than the values derived by Velev et al. (22). This result was obtained by using a different procedure for fitting the scattering data and the use of very low protein concentration solutions (24). All computed and experimental values are given in Table S2.

BPTI

Simulations were performed at 300, 500, 700, and 800 mM IS at pH 4.9, following the experiments of Farnum and Zukoski (47). The protein concentration was set to 20 g/L, which is higher than the experimental one (3 g/L) but enables quicker convergence of the $g(r)$ with simulation time (see the Supporting Material). Experimentally, it was found that at 300 mM IS, the B_{22} value is positive (47,48), meaning that the interactions are, on average, repulsive. At 500 mM IS, the B_{22} value is ≈ 0, and at higher IS values, the B_{22} values are negative. This trend is reproduced qualitatively by the computations (see Fig. 2). As for HEWL, in Fig. 2 (and in Table S2) we also added the long-range analytical correction to the computed B_{22} values, which results in an overestimation of the lower ionic strength values. Analogous to the results of Asthagiri et al. (24) for HEWL, we may expect that the experimental values are underestimated.

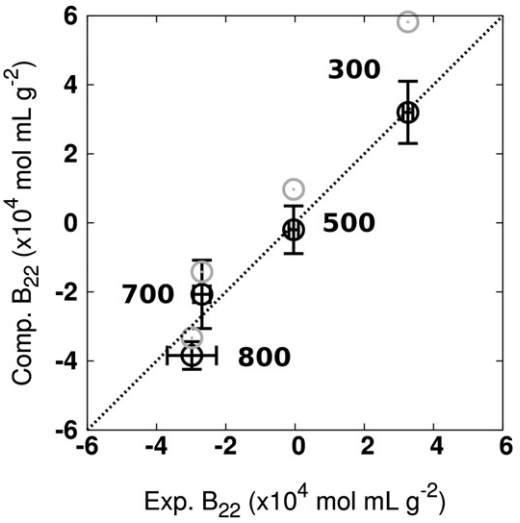

FIGURE 2 Plot of computed against experimental (47) B_{22} values for BPTI at four different IS (mM) shown by labels. Each point corresponds to a different IS (mM) as labeled on the plot. A reference line with gradient 1 is dashed. The computed values are given with and without the analytical long-range correction as shaded and solid symbols, respectively.

Net attractive interactions occur only above an IS of 500 mM at which the electrostatic repulsion is strongly screened (47). Indeed, checking the average value of the electrostatic interaction energy between one protein and all the others in the simulations, we found that at 300 mM IS, it is slightly positive at 0.002 ± 0.289 kT, whereas at IS $\geq$ 500 mM, it is -0.03 ± 0.301 kT.

T4 lysozyme

Theoretical models, such as the DLVO model, and simplified computational models which do not take the atomic details of the interacting macromolecules into account, are very valuable for providing a phenomenological description of the behavior of structural properties such as B_{22} values. On the other hand, they cannot provide any structural interpretation of such properties (10). Moreover, changes in the magnitude or sign of B_{22} values due to single point mutations which do not affect the net charge of the protein, or which affect the shape of the molecule, cannot be described by simplified models.

To test the ability of our model to describe effects due to point mutations, we, following McGuffee and Elcock (11), compared wild-type (WT) T4 lysozyme with three single point mutants (S44K, S44F, and S44E) whose B_{22} values have been measured (49). McGuffee and Elcock (11) found that it was possible to qualitatively reproduce the effect of mutations at residue 44 on the B_{22} values from simulations.

Whereas the introduction of lysine or glutamic acid at position 44 affects steric and electrostatic properties, the S44F mutant has a bulky aromatic residue which mainly influences steric interactions without perturbing the electrostatic interactions significantly. This implies that nonpolar interactions and steric effects, which cannot be reproduced by simplified models, become important.

Simulations were done at 50 mM IS, pH 7, and 20 g/L protein concentration to reproduce the experimental conditions (49). The computed B_{22} values reproduce the trends observed experimentally and agree within the experimental confidence interval (see Fig. 3). For the WT protein, the computed B_{22} value is negative. For the S44K mutant, the experimental B_{22} value is higher than for the WT, implying greater repulsion between the proteins (49). This behavior is consistent with the increase in the net charge of the protein from $+8e$ for the WT to $+9e$ for the S44K mutant, and is reproduced computationally.

For the S44F mutant, the B_{22} value was found to be negative experimentally, implying an attractive interaction between the proteins (49). The computed B_{22} value captures the change of interaction strength between the WT and the S44F mutant. To determine whether the increased attraction arises from formation of a nonpolar contact patch, the orientation of the monomers in the dimers found in the simulations was explored (see Oligomer Formation, below).

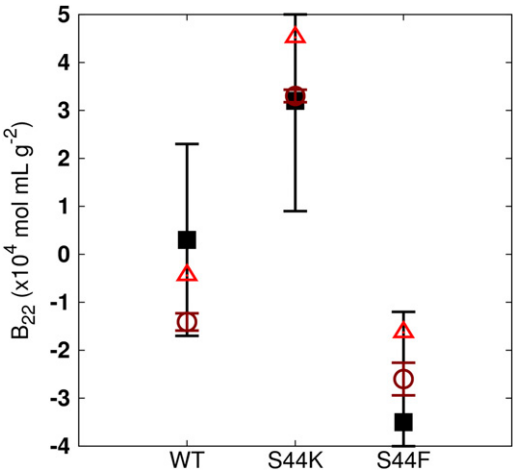

FIGURE 3 Experimental (49) (*black squares*) and computed (*circles and triangles*) B_{22} values for WT T4 lysozyme and the S44K and S44F mutants. The computed values are given with and without the analytical long-range correction as light-red triangles and dark-red circles, respectively.

For the last mutant studied, S44E, it was not possible to obtain a reliable B_{22} value experimentally because of precipitation of the mutant during purification (49). The mutation reduces the net charge at pH 7 from $+8e$ in the WT to $+7e$ in the S44E mutant. The computed B_{22} value

$$-1.96 \pm 0.3 \left(\times 10^4 \text{ mL mol/g}^2 \right)$$

and

$$-2.7 \pm 0.3 \left(\times 10^4 \text{ mL mol/g}^2 \right)$$

with and without the analytical correction, respectively, for the S44E mutant is more negative than for WT, indicating increased attraction which may be due to diminished electrostatic repulsion between proteins.

Oligomer formation

The fractions of proteins in different oligomeric states were computed as described in Methods. The fraction of monomers is given by χ_1, whereas the fraction of n-mers is χ_n.

HEWL

The aggregation and equilibrium cluster formation properties of HEWL solutions are the subject of considerable debate (50–52). In a recent study based on neutron spin echo and small angle neutron scattering, Porcar et al. (52) found that, at low concentrations of lysozyme (V = [0.05–0.1]), solutions consist mostly of monomers or dimers, whereas, at high concentrations (V = [0.15–0.2]), large dynamic clusters are dominant. Porcar et al. interpreted the cluster peak in the SANS/SAXS pattern as due to transient clusters, whose shape and aggregation number fluctuate. These findings contradict previous conclusions from Shukla et al. (51), who claimed that the system contains largely repulsive individual lysozyme proteins. Our simulations were performed at 14, 28, 42, and 57 g/L protein concentration, at 150 mM IS and pH 4.6 for comparison was made to experimental values from Price et al. (53).

In this concentration range, in agreement with Porcar et al. (52), we found the solution mainly consists of monomers (70–90%) and dimers (10–20%) (see Fig. 4 *B*). We also tested the effect of ionic strength on the cluster formation. At 5 mM IS, electrostatic repulsion is large and the fraction of monomers was 100% in the simulations. On increasing IS, the fraction of dimers increased due to the decrease of repulsive electrostatic interactions. At 100 mM IS, the fraction of monomers reduced to 91% and an 8% fraction of dimers appeared. This change in oligomeric state can be monitored by comparing the computed structure factors ($S(q)$) with the experimental $S(q)$ obtained by small angle scattering (51). Although the simulations were not performed under exactly the same conditions as these experiments, they reproduce a similar dependence of the interference peak at low q values on IS (see Fig. 4 *A*), which is consistent with a decrease in repulsive interactions and a slightly increased attraction (51).

BPTI

The self-association of BPTI has been debated for a long time based on SAS and DLS (54–58), NMR pulsed-gradient spin-echo (59), crystallographic (57,60,61), and magnetic relaxation dispersion (61) experiments. Crystal growth at high pH (9,10), close to the isoelectric point, shows one BPTI molecule per asymmetric unit (62,63), but more recent studies at lower pH values (4.5–7) show crystal forms with five or 10 BPTI molecules in the asymmetric unit, representing different stackings of a common decamer structure (57,60). At pH 4.9, BPTI solutions were found to contain

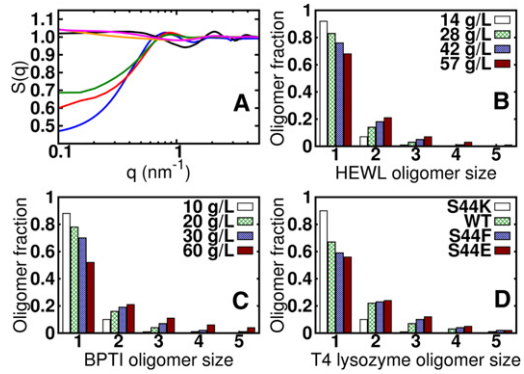

FIGURE 4 (*A*) Experimental structure factor, $S(q)$, at 22 g/L HEWL, 50 mM IS, pH 7.8 (*black*) (51), is compared with the computed $S(q)$ at 10 g/L HEWL, 100 mM IS, pH 6 (*orange*) and 9 (*magenta*). The experimental $S(q)$ at 15 g/L HEWL, 6 mM IS, pH 7.8 (*blue*), is compared to the computed $S(q)$ at 10 g/L HEWL, 5 mM IS, pH 6 (*red*) and 9 (*green*). (*B–D*) Fractions of monomer χ_1 and higher oligomeric states χ_n observed in HEWL (*B*), BPTI (*C*), and T4 lysozyme (*D*) simulations. A contact was defined by the parameters (see Methods): $N = 2$, $d_{min} = 6.0$ Å, and $d_c = 4.5$ Å.

only monomers and decamers, and the decamer was proposed to represent the growth unit for the crystal (57,61).

We performed simulations at 300 mM IS and pH 7, and at 10, 20, 30, and 60 g/L protein concentrations, according to Tanaka et al. (64). The fraction of oligomers was higher than in the HEWL solutions. Over this concentration range,

$$\chi_1 = [0.88 - 0.52] \text{ and } \chi_2 = [0.10 - 0.21].$$

Moreover, in contrast to HEWL solutions, some higher order oligomers were detected with

$$\chi_3 = 0.11,\ \chi_4 = 0.06, \text{ and } \chi_5 = 0.04 \text{ at } 60 \text{ g/L}$$

(see Fig. 4 *C* and Fig. S2). However, no symmetric pentamers or decamers of the type found in the low pH crystal structures were observed during the simulations. This is not surprising, when one considers that formation of the decamer in the crystal relies on the specific binding of anions which alleviates the electrostatic repulsion between basic residues in the channel inside the decamer (57,61). This specific anion binding is inadequately described by a continuum solvent model with Debye screening.

A cluster analysis of the dimers identified in our simulations reveals that the two monomers have complementary electrostatic interactions. A positive patch in the region of the loop 13-18 (the essential loop for the inhibition of the bovine trypsin) interacts with a slightly negative region close to the C-terminus. A similar interaction is observed in the crystal contacts in several monomeric crystal structures of BPTI. However, the dimers formed in the simulations do not correspond to the A-G dimer formed across the twofold axis in the crystal structure which was proposed to be the first dimer formed in the process of decamer assembly (57). In the crystal structure, the A-G dimer interface has a carboxy oxygen of D50 of monomer A 2.5 Å from a carboxy oxygen of D50 of monomer G (61). Their interaction relies on a change in protonation state due to a shift in pK_a (61), which is facilitated at low pH. In our simulations, the distance between the oxygen atoms of the two aspartic acids is always >4 Å and the monomers cannot orient to allow the formation of the hydrogen-bond network that stabilizes the dimer in the crystal structure.

Comparison of the experimental data and the simulation data support the idea that decamer assembly is a highly cooperative process that requires the presence of anions and a low pH. Indeed, in general, crystallization and nucleation processes have high sensitivity to environmental conditions, and to impurities and contaminants that can be present in solution, making them nontrivial phenomena for both experimental and computational studies (65–67).

T4 lysozyme

Computed oligomer fractions are shown in Fig. 4 *D*. In agreement with the B_{22} values, the S44K mutant is mainly monomeric in solution ($\chi_1 \approx 0.90$). The S44F and S44E mutants show similar behavior to each other, with $\chi_1 \approx 0.50$ and $\chi_2 \approx 0.25$. A slightly higher fraction of dimers and higher order oligomers is found for the S44E mutant compared to the S44F mutant, in accord with its lower B_{22} value.

To assess the importance of the hydrophobic patch in the S44F mutant (49), a cluster analysis (see Methods) on the identified dimers was performed for the S44F mutant and for the WT. The first two of the five clusters identified for the WT each represent 35% of the dimers. For the S44F mutant, six clusters were identified. Each of the first four clusters represents ~20% of the dimers. The two most populated clusters of the WT (Fig. 4, *A* and *B*) and the S44F mutant (Fig. 4, *C* and *D*) are compared in Fig. 5. For the WT, S44 is not involved in interactions with other proteins in these two clusters. Both orientations are favored by electrostatic interactions between oppositely charged patches (Fig. S3). These contacts do not reproduce the crystal contacts found in the x-ray structure.

For S44F, F44 is directly involved in an interprotein contact in these two clusters. Moreover, mapping the hydrophobicity on the van der Waals surface of one monomer, we found that the interaction occurs through hydrophobic patches, confirming the importance of the hydrophobic patch hypothesis proposed by Chang et al. (49). From the computed atomic contact probabilities, it can be seen that F44 is among the residues with the highest contact probability (see Fig. 6 *B*), in agreement with the result of McGuffee and Elcock (11).

For the S44E mutant, the interaction was found to occur mainly via E44, which shows a high contact probability (see Fig. 6 *C*).

The S44K mutant, in agreement with its positive B_{22} value, has a very low propensity to interact, as shown by

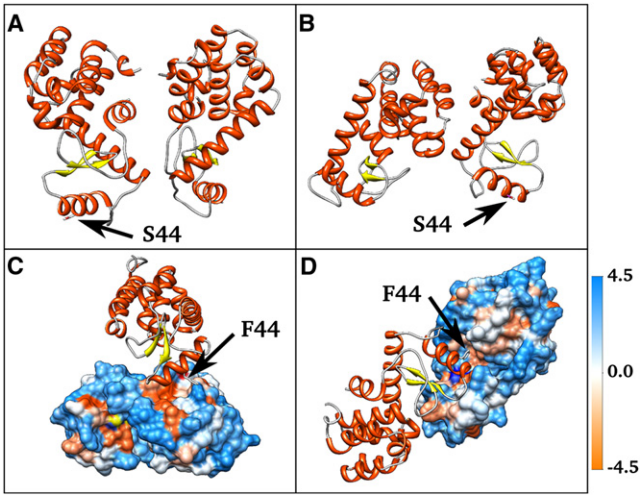

FIGURE 5 Most common arrangements of T4 lysozyme monomers in dimers obtained from clustering analysis of simulation snapshots: WT: cluster 1 (*A*) and 2 (*B*); S44F mutant: cluster 1 (*C*) and 2 (*D*).

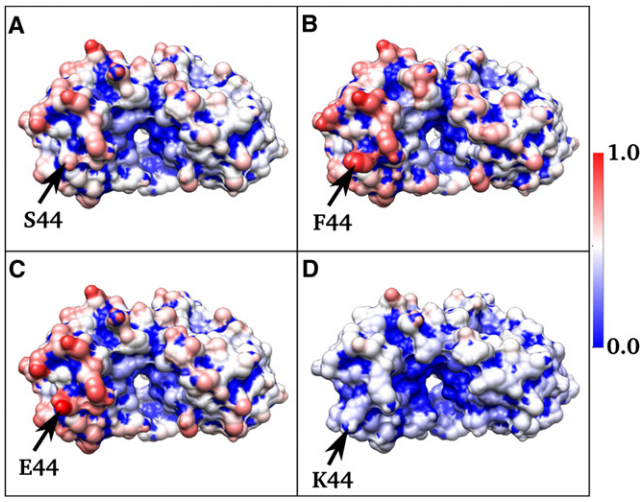

FIGURE 6 Relative atomic contact probabilities are displayed on the van der Waals surface of T4 lysozyme. The WT (*A*) and the S44F (*B*), S44E (*C*), and S44K (*D*) mutants are shown.

the mapping of atomic contact probabilities on its structure displayed in Fig. 6 *D*.

Self-diffusion coefficients (D_s)

HEWL

The computed self-diffusion coefficients agree well with the experimental results (see Fig. 7). Theoretical models that describe the volume fraction dependence of the self-diffusion coefficient were compared to the simulation results. The Han and Herzfeld model (27), which is based on scaled particle theory, is not able to correctly reproduce the decrease of the self-diffusion coefficient with increasing concentration. Using a more sophisticated model developed by Tokuyama and Oppenheim (29), which takes hydrodynamic interactions into account, the volume fraction dependence of the diffusion coefficient can be correctly described.

Neither of these theoretical models takes the formation of oligomers into account. This implies that the reduction of the self-diffusion coefficient of HEWL at intermediate IS can be described in terms of hindered diffusion arising mostly from an excluded volume effect and hydrodynamic interactions. These models cannot, however, be applied when a sizeable number of dimers or higher order oligomers form in the solutions (see next section on BPTI).

BPTI

The computed self-diffusion coefficients were compared to experimental values from Tanaka et al. (64), which agree with previously reported values from Gallagher and Woodward (54). Both sets of diffusion coefficients were obtained by DLS. The diffusion coefficient obtained using DLS is the collective or mutual diffusion coefficient (D_m), which is related to the self-diffusion coefficient via the second virial coefficient (46,68) (see the Supporting Material).

The experimental second virial coefficient at 20 g/L and 300 mM IS is 3.3 ($\times 10^4$ mL mol/g^2). Considering a partial specific volume of BPTI of $v_{sp} = 0.71$ cm^3/g (69), the ratio is

$$\delta = \frac{D_m}{D_s} = 1.03.$$

Hence, the computed self-diffusion coefficients can be directly compared to the rescaled experimental values. A good agreement between experimental and computed diffusion coefficients was found, as shown in Fig. 8.

It is worth noting that the Han and Herzfeld (27) and Tokuyama and Oppenheim (29) theoretical models fail in the description of the concentration-dependent behavior of the diffusion coefficient. As previously mentioned, this is related to the lack of a description of oligomer formation

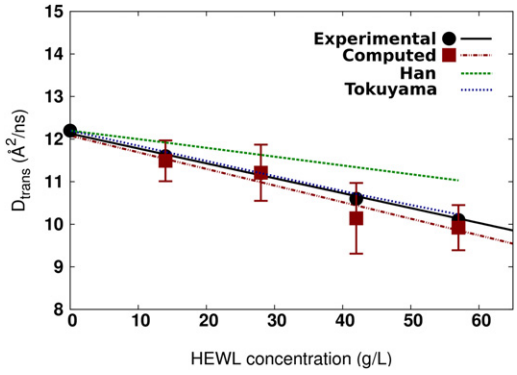

FIGURE 7 Dependence of the self-diffusion coefficients of HEWL on protein concentration (150 mM IS, pH 4.5). Experimental values (53) are shown by black circles, computed values by dark-red squares. Linear regression lines for experimental and computed values are depicted by a black-solid line and dark-red dot-dash line, respectively. The Han and Herzfeld (27) and Tokuyama and Oppenheim (29) theoretical models are shown in green-dashed and blue-dotted lines, respectively.

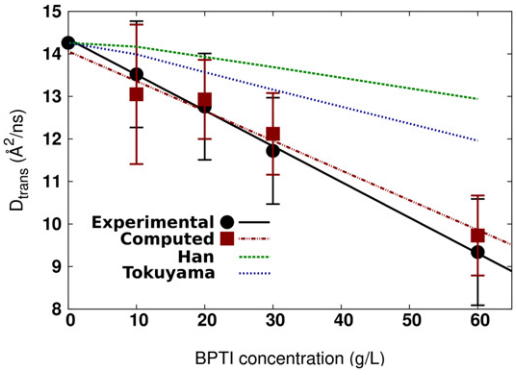

FIGURE 8 Dependence of the self-diffusion coefficients of BPTI on protein concentration (300 mM IS, pH 7). Experimental values (64) are shown by black circles, computed values by dark-red squares. Linear regression lines for experimental and computed values are depicted by a black-solid line and dark-red dot-dash line, respectively. The Han and Herzfeld (27) and Tokuyama and Oppenheim (29) theoretical models are shown in green-dashed and blue-dotted lines, respectively.

in these models. Particularly for the highest concentration, a significant fraction of dimers and some high order oligomers is formed (Fig. 4 *C*). Hence, the excluded volume effect and the hydrodynamic interactions account only partially for the reduction in the diffusion coefficient, and aggregation must be considered.

CONCLUSIONS

In summary, we have described a BD method for simulating many atomically detailed rigid macromolecules in implicit solvent. The model of the forces between macromolecules is generally applicable to the simulation of proteins in aqueous solution. For the tested systems, a good accord between the calculated values and the experimental thermodynamic (second virial coefficients) and dynamic (translational self-diffusion coefficients) properties was found.

The principal limitation of the simulation model is the neglect of the internal flexibility of the macromolecules. This limits application to association processes that do not involve large changes in macromolecular conformation.

The effects due to hydrodynamic interactions are not as significant as in simulations of flexible polymers. Therefore, at the low volume fractions used in this work, they can be neglected (11,70). Test calculations of the diffusion coefficient with inclusion of hydrodynamic interactions in the model showed no significant difference in the computed values (P Mereghetti and R Wade, unpublished data).

Previous simulations of systems of many diffusing proteins describe the interactions using different levels of approximation, e.g., with proteins as spherical particles, a mixture of sphere and atomic-detail representations, or in atomic detail (11,13,44,71,72). The atomic-detail methodology of McGuffee and Elcock (11,13) is most similar to ours but differs primarily in the modeling of the intermolecular interactions. McGuffee and Elcock (11,13) used an electrostatic term and an adjustable hydrophobic term (Lennard-Jones potential). The latter was fit separately for each protein studied to reproduce experimental thermodynamic data.

In our model, desolvation effects are included by introducing electrostatic and nonpolar desolvation terms. Moreover, none of the parameters of our model are fit to any specific system and thus, the approach may be of value for the prediction of static and dynamic properties of solutions of macromolecules, at a level that is sensitive to alteration of single amino-acid residues. For example, this approach can be used to study cooperative aggregation or polymerization processes as well as macromolecular diffusion and association kinetics.

SUPPORTING MATERIAL

Seven equations, two tables, three figures, and text describing the computational details of the simulations and their analysis are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)01323-8.

This work was supported by the Klaus Tschira Foundation and the Center for Modeling and Simulation in the Biosciences in Heidelberg, and by supercomputing time at the Environmental Molecular Science Laboratory (grant No. 30994).

REFERENCES

1. Minton, A. P. 2001. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* 276:10577–10580.
2. Ellis, R. J. 2001. Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* 26:597–604.
3. Homouz, D., L. Stagg, …, M. S. Cheung. 2009. Macromolecular crowding modulates folding mechanism of α/β protein apoflavodoxin. *Biophys. J.* 96:671–680.
4. Zhou, H.-X., G. Rivas, and A. P. Minton. 2008. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* 37:375–397.
5. Kim, J. S., and A. Yethiraj. 2009. Effect of macromolecular crowding on reaction rates: a computational and theoretical study. *Biophys. J.* 96:1333–1340.
6. Batra, J., K. Xu, …, H. X. Zhou. 2009. Effect of macromolecular crowding on protein binding stability: modest stabilization and significant biological consequences. *Biophys. J.* 97:906–911.
7. Mittal, J., and R. B. Best. 2010. Dependence of protein folding stability and dynamics on the density and composition of macromolecular crowders. *Biophys. J.* 98:315–320.
8. Huppa, J. B., M. Axmann, …, M. M. Davis. 2010. TCR-peptide-MHC interactions in situ show accelerated kinetics and increased affinity. *Nature.* 463:963–967.
9. Batra, J., K. Xu, and H.-X. Zhou. 2009. Nonadditive effects of mixed crowding on protein stability. *Proteins.* 77:133–138.
10. Elcock, A. H., D. Sept, and J. A. McCammon. 2001. Computer simulation of protein-protein interactions. *J. Phys. Chem. B.* 105:1504–1518.
11. McGuffee, S. R., and A. H. Elcock. 2006. Atomically detailed simulations of concentrated protein solutions: the effects of salt, pH, point mutations, and protein concentration in simulations of 1000-molecule systems. *J. Am. Chem. Soc.* 128:12098–12110.
12. Klein, M. L., and W. Shinoda. 2008. Large-scale molecular dynamics simulations of self-assembling systems. *Science.* 321:798–800.
13. McGuffee, S. R., and A. H. Elcock. 2010. Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLOS Comput. Biol.* 6:e1000694.
14. Gabdoulline, R. R., and R. C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* 72:1917–1929.
15. Elcock, A. H. 2010. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol.* 20:196–206.
16. Gabdoulline, R. R., and R. C. Wade. 2009. On the contributions of diffusion and thermal activation to electron transfer between *Phormidium laminosum* plastocyanin and cytochrome *f*: Brownian dynamics simulations with explicit modeling of nonpolar desolvation interactions and electron transfer events. *J. Am. Chem. Soc.* 131:9230–9238.
17. Gabdoulline, R. R., and R. C. Wade. 1996. Effective charges for macromolecules in solvent. *J. Phys. Chem.* 100:3868–3878.
18. Elcock, A. H., R. R. Gabdoulline, …, J. A. McCammon. 1999. Computer simulation of protein-protein association kinetics: acetylcholinesterase-fasciculin. *J. Mol. Biol.* 291:149–162.
19. Gabdoulline, R. R., and R. C. Wade. 1998. Brownian dynamics simulation of protein-protein diffusional encounter. *Methods.* 14:329–341.
20. Madura, J. D., J. M. Briggs, …, J. A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* 91:57–95.

21. McMillan, W. G., and J. E. Mayer. 1945. The statistical thermodynamics of multicomponent systems. *J. Chem. Phys.* 13:276–305.
22. Velev, O. D., E. W. Kaler, and A. M. Lenhoff. 1998. Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen. *Biophys. J.* 75:2682–2697.
23. Zimm, B. H. 1948. The scattering of light and the radial distribution function of high polymer solutions. *J. Chem. Phys.* 16:1093–1099.
24. Asthagiri, D., A. Paliwal, …, M. E. Paulaitis. 2005. A consistent experimental and modeling approach to light-scattering studies of protein-protein interactions in solution. *Biophys. J.* 88:3300–3309.
25. Hill, T. L. 1986. An Introduction to Statistical Thermodynamics. Dover, Mineola, NY.
26. Allen, M. P., and D. J. Tildesley. 1989. Computer Simulations of Liquids. Oxford University Press, New York.
27. Han, J., and J. Herzfeld. 1993. Macromolecular diffusion in crowded solutions. *Biophys. J.* 65:1155–1161.
28. Reiss, H., H. L. Frisch, and J. L. Lebowitz. 1959. Statistical mechanics of rigid spheres. *J. Chem. Phys.* 31:369–380.
29. Tokuyama, M., and I. Oppenheim, I. 1994. Dynamics of hard-sphere suspensions. *Phys. Rev. E.* 50:R16–R19.
30. Berman, H. M., J. Westbrook, …, P. E. Bourne. 2000. The Protein DataBank. *Nucleic Acids Res.* 28:235–242.
31. Wilson, K. P., B. A. Malcolm, and B. W. Matthews. 1992. Structural and thermodynamic analysis of compensating mutations within the core of chicken egg white lysozyme. *J. Biol. Chem.* 267:10842–10849.
32. Lubkowski, J., and A. Wlodawer. 1999. Decamers observed in the crystals of bovine pancreatic trypsin inhibitor. *Acta Crystallogr. D Biol. Crystallogr.* 55:335–337.
33. Eriksson, A. E., W. A. Baase, and B. W. Matthews. 1993. Similar hydrophobic replacements of Leu^{99} and Phe^{153} within the core of T4 lysozyme have different structural and thermodynamic consequences. *J. Mol. Biol.* 229:747–769.
34. Gordon, J. C., J. B. Myers, …, A. Onufriev. 2005. HPP: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Res.* 33:368–371.
35. Abe, Y., T. Ueda, …, T. Imoto. 1995. Effect of salt concentration on the pK_a of acidic residues in lysozyme. *J. Biochem.* 118:946–952.
36. Demchuk, E., and R. C. Wade. 1996. Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins. *J. Phys. Chem.* 100:17373–17387.
37. Kuhn, B., P. A. Kollman, and M. Stahl. 2004. Prediction of pK_a shifts in proteins using a combination of molecular mechanical and continuum solvent calculations. *J. Comput. Chem.* 25:1865–1872.
38. Kuehner, D. E., J. Engmann, …, J. M. Prausnitz. 1999. Lysozyme net charge and ion binding in concentrated aqueous electrolyte solutions. *J. Phys. Chem. B.* 103:1368–1374.
39. Jorgensen, W. L., and J. Tirado-Rives. 1988. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* 110:1657–1666.
40. Sali, A., and T. L. Blundell. 1993. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* 234:779–815.
41. McQuarrie, D. A. 1976. Statistical Mechanics. Harper and Row, New York.
42. Neal, B. L., D. Asthagiri, and A. M. Lenhoff. 1998. Molecular origins of osmotic second virial coefficients of proteins. *Biophys. J.* 75:2469–2477.
43. Elcock, A. H., and J. A. McCammon. 2001. Calculation of weak protein-protein interactions: the pH dependence of the second virial coefficient. *Biophys. J.* 80:613–625.
44. Carlsson, F., M. Malmsten, and P. Linse. 2001. Monte Carlo simulations of lysozyme self-association in aqueous solution. *J. Phys. Chem. B.* 105:12189–12195.
45. Lund, M., and B. Jönsson. 2003. A mesoscopic model for protein-protein interactions in solution. *Biophys. J.* 85:2940–2947.
46. Teraoka, I. 2002. Polymer Solutions: An Introduction to Physical Properties. Wiley, New York.
47. Farnum, M., and C. Zukoski. 1999. Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor. *Biophys. J.* 76:2716–2726.
48. Appavou, M. S., G. Gibrat, and M.-C. Bellissent-Funel. 2009. Temperature dependence on structure and dynamics of Bovine Pancreatic Trypsin Inhibitor (BPTI): a neutron scattering study. *Biochim. Biophys. Acta.* 1794:1398–1406.
49. Chang, R. C., D. Asthagiri, and A. M. Lenhoff. 2000. Measured and calculated effects of mutations in bacteriophage T4 lysozyme on interactions in solution. *Proteins.* 41:123–132.
50. Stradner, A., F. Cardinaux, and P. Schurtenberger. 2006. A small-angle scattering study on equilibrium clusters in lysozyme solutions. *J. Phys. Chem. B.* 110:21222–21231.
51. Shukla, A., E. Mylonas, …, D. I. Svergun. 2008. Absence of equilibrium cluster phase in concentrated lysozyme solutions. *Proc. Natl. Acad. Sci. USA.* 105:5075–5080.
52. Porcar, L., P. Falus, …, Y. Liu. 2009. Formation of the dynamic clusters in concentrated lysozyme protein solutions. *J. Phys. Chem. Lett.* 1: 126–129.
53. Price, W. S., F. Tsuchiya, and Y. Arata. 1999. Lysozyme aggregation and solution properties studied using PGSE NMR diffusion measurements. *J. Am. Chem. Soc.* 121:11503–11512.
54. Gallagher, W. H., and C. K. Woodward. 1989. The concentration dependence of the diffusion coefficient for bovine pancreatic trypsin inhibitor: a dynamic light scattering study of a small protein. *Biopolymers.* 28:2001–2024.
55. Veesler, S., S. Lafont, …, R. Boistelle. 1996. Prenucleation, crystal growth and polymorphism of some proteins. *J. Cryst. Growth.* 168:124–129.
56. Lafont, S., S. Veesler, …, R. Boistelle. 1997. Comparison of solubilities and molecular interactions of BPTI molecules giving different polymorphs. *J. Cryst. Growth.* 173:132–140.
57. Hamiaux, C., J. Pérez, …, P. Vachette. 2000. The BPTI decamer observed in acidic pH crystal forms pre-exists as a stable species in solution. *J. Mol. Biol.* 297:697–712.
58. Grouazel, S., F. Bonneté, …, S. Veesler. 2006. Exploring bovine pancreatic trypsin inhibitor phase transitions. *J. Phys. Chem. B.* 110:19664–19670.
59. Ilyina, E., V. Roongta, …, K. H. Mayo. 1997. A pulsed-field gradient NMR study of bovine pancreatic trypsin inhibitor self-association. *Biochemistry.* 36:3383–3388.
60. Hamiaux, C., T. Prangé, …, S. Veesler. 1999. The decameric structure of bovine pancreatic trypsin inhibitor (BPTI) crystallized from thiocyanate at 2.7 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* 55:103–113.
61. Gottschalk, M., and B. Halle. 2003. Self-association of lysozyme as seen by magnetic relaxation dispersion. *J. Phys. Chem. B.* 107:7914–7922.
62. Deisenhofer, J., and W. Steigemann. 1975. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Crystallogr. B.* 31:238–250.
63. Wlodawer, A., J. Walter, …, L. Sjölin. 1984. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and x-ray refinement of crystal form II. *J. Mol. Biol.* 180:301–329.
64. Tanaka, S., M. Ataka, …, S. Veesler. 2002. pH-dependent oligomerization of BPTI in undersaturated and supersaturated solutions studied by dynamic light scattering. *J. Cryst. Growth.* 237:289–294.
65. Vekilov, P. G., B. R. Monaco, …, F. Rosenberger. 1996. Repartitioning of NaCl and protein impurities in lysozyme crystallization. *Acta Crystallogr. D Biol. Crystallogr.* 52:785–798.
66. Thomas, B. R., P. G. Vekilov, and F. Rosenberger. 1998. Effects of microheterogeneity in hen egg-white lysozyme crystallization. *Acta Crystallogr. D Biol. Crystallogr.* 54:226–236.
67. Forsythe, E. L., A. Nadarajah, and M. L. Pusey. 1999. Growth of (101) faces of tetragonal lysozyme crystals: measured growth-rate trends. *Acta Crystallogr. D Biol. Crystallogr.* 55:1005–1011.

68. Zhang, J., and X. Y. Liu. 2003. Effect of protein-protein interactions on protein aggregation kinetics. *J. Chem. Phys.* 119:10972–10976.

69. Makhatadze, G. I., K. S. Kim, …, P. L. Privalov. 1993. Thermodynamics of BPTI folding. *Protein Sci.* 2:2028–2036.

70. Antosiewicz, J., and J. A. McCammon. 1995. Electrostatic and hydrodynamic orientational steering effects in enzyme-substrate association. *Biophys. J.* 69:57–65.

71. Wieczorek, G., and P. Zielenkiewicz. 2008. Influence of macromolecular crowding on protein-protein association rates—a Brownian dynamics study. *Biophys. J.* 95:5030–5036.

72. Li, X., I. H. Moal, and P. A. Bates. 2010. Detection and refinement of encounter complexes for protein-protein docking: taking account of macromolecular crowding. *Proteins: Struct. Funct. Bioinform.* 78:3189–3196.